A13073W1

DEGREE OF MASTER OF SCIENCE IN FINANCIAL ECONOMICS

FINANCIAL ECONOMETRICS

TRINITY TERM 2020

Tuesday, 21 April 2020

Time allowed is FOUR HOURS

You **MUST** upload your submission within 4 hours of accessing the paper

*Candidates should answer **ALL** questions in part A.*
*Candidates should answer **THREE** of questions in part B.*
*Candidates should answer **ALL** questions in part C.*

Examiners will place weight 2% on each question in Part A (40% total),
10% on each question in part B and 30% on Part C.

*Please use **the solutions template** provided if possible.*

Materials: Candidates may use their any calculator or any other software when preparing their answer.

Do not turn over until told that you may do so.

PART A: MULTIPLE CHOICE

Answer ALL questions in this section.

The section contributes 40% towards the final mark. Each question is worth 2% of the exam mark.

Single Answer Questions

Select a single answer for each question. Correct marks are awarded 2%, and incorrect marks reduce the score by 0.5% so that random guessing does not improve your expected mark.

1. If you flip three fair coins, what is the probability that all three show the same side?

   (a) $1/2$

   (b) $1/8$

   (c) $1/16$

   (d) $1/4$

2. When evaluating a series of forecasts using the Mincer-Zarnowitz regression $y_{t+h} - \hat{y}_{t+h|t} = \alpha + \beta \hat{y}_{t+h|t} + \eta_{t+h}$, what are the values of α and β that should occur when the forecasting model is correctly specified.

   (a) $\beta = 0$, no restriction on α

   (b) $\alpha = 0, \beta = 0$

   (c) $\alpha = 0, \beta = 1$

   (d) $\beta = 1$, no restriction on α

3. Daily return on a portfolio are i.i.d. with a $N\left(0.05\%, (1\%)^2\right)$ distribution. What is the 1-week 5% Value-at-Risk or a portfolio with £1,000,000,000 under management (to the nearest 10,000)?

   (a) 2,500,000

   (b) 34,280,000

   (c) 41,330,000

   (d) 4,130,000

4. How are Diebold-Mariano tests used to compare the forecasts from two VaR models?

(a) The two series of HITs from the models are used as losses.

(b) HITs are regressed on lagged HITs and the two VaR forecasts.

(c) Diebold-Mariano tests cannot be used to compare two VaR models.

(d) The two series of HITs are transformed using the tick-loss function, and then the difference is tested.

5. How are simulated values used in Historical Simulation VaR?

(a) They are not. The quantile depends only on α and the sample size n.

(b) They rely on simulated models produced in the past.

(c) Multi-step VaR uses values simulated from historical observations.

(d) Returns are first filtered by an ARCH-type model and then simulated forward to estimate the VaR.

6. What restrictions are required on Φ_1 for a VAR(1)

$$y_t = \Phi_1 y_{t-1} + \varepsilon_t$$

to be stationary?

(a) All eigenvalues must be positive.

(b) All eigenvalues must be less than 1 in modulus.

(c) VAR(1) models are always covariance stationary. Only VAR(P) models can be non-stationary.

(d) All values in Φ_1 must be less than 1 in absolute value.

7. The Local in Local Average Treatment Effects refers to:

(a) The estimate of the ATE is weighted to reflect the probability of participation in a randomized controlled trial (RCT).

(b) The reality the all experiments only use subjects in one locality.

(c) That the maximum likelihood estimator of the ATE may achieve a local rather than a global maximum.

(d) The effect is homogeneous across the treatment group.

8. Consider the model $y_t = y_{t-1} + \varepsilon_t$, where $\varepsilon_t \overset{\text{i.i.d.}}{\sim} N(0, \sigma_\varepsilon^2)$. Which statement below false?

(a) $\mathrm{E}[y_t]$ grows over time.

(b) A regression of y_t on x_t with $x_t = 0.5x_{t-1} + v_t$ where $v_t \sim N(0, \sigma_v^2)$ and where ε_t and v_t are independent can result in spurious correlation.

(c) There is no mean reversion in long-run forecasts of y_t so that $\mathrm{E}_t\left[y_{t+h}\right] = y_t$ for any $h \geq 1$.

(d) In a regression $y_t = \beta y_{t-1} + \varepsilon_t$, our OLS estimate $\hat{\beta}$ would be inconsistent.

9. How does the GJR-GARCH model improve on the GARCH model?

   (a) It models the log-variance instead of the variance, and so is always positive.

   (b) It allows for more lags of the squared return and variance.

   (c) It adds an asymmetry term that depends on the sign of the past return.

   (d) It models the standard deviation instead of the variance.

10. In a hypothesis test, the power of the test is

   (a) The probability the null is rejected given the alternative is true.

   (b) the probability of a Type I error.

   (c) the probability of a Type II error.

   (d) The probability that the alternative is true.

Multiple Answer Questions

Select all correct answers. Each question has between 0 and n correct answers where n is the number of options in the questions. Each of the n answers is treated as a true-false question so that a correct subpart answer is awarded $^{2\%}/_{n}$, and an incorrect answer reduces the mark by $^{2\%}/_{n}$. For example, if the correct answers in a 5-part question are A, D, and E, then an answer with A is awarded 0.4%, and an answer without A is reduced by 0.4%. An answer with B is reduced by 0.4%, and an answer without B is awarded 0.4%. Random guessing does not improve your expected mark.

11. If $\mathrm{E}[Y|X] = 0$, where X is uniformly distributed on $[5, 10]$, then which of the following statements are true:

   (a) $\mathrm{E}\left[g\left(Y\right)|X\right] = 0$ for any well defined function $g\left(\cdot\right)$

   (b) $E[YX] = 0$

   (c) $\mathrm{E}[^{Y}/_{X}] = 0$

   (d) $\mathrm{Cov}\left[X, Y\right] = 0$

12. What restrictions are required in an APARCH model to get an ARCH(1)?

$$\sigma_t^{\delta} = \omega + \alpha\left(|\varepsilon_{t-1}| + \gamma\varepsilon_{t-1}\right)^{\delta} + \beta\sigma_{t-1}^{\delta}$$

   (a) γ=0

   (b) $\beta = 0$

(c) $\delta = 0$

(d) $\beta = 1 - \alpha$

(e) $\delta = 2$

13. Why is the companion form of a VAR(P) useful?

(a) It allows any VAR(P) to be expressed as a VAR(1)

(b) It simplifies computing the autocovariance function of a VAR.

(c) At allows an AR(P) to be written as a VAR(1)

(d) It transforms a VAR to ensure that it is covariance stationary.

14. Which of the following are true about the expectations operator, $\mathrm{E}[\cdot]$:

(a) $\mathrm{E}[XY] = \mathrm{Cov}[X,Y]$ only when the random variables X and Y are independent

(b) If $g(\cdot)$ is a concave function, then $\mathrm{E}[g(X)] \geq g[\mathrm{E}[X]]$

(c) $\mathrm{V}[a+bX] = b\mathrm{V}[X]$ where a and b are constants and X is a random variable

(d) $\mathrm{E}[\mathrm{E}[X|Y]] = \mathrm{E}[X]$ only if X and Y are independent

15. Which are true of a vector white noise process $\{\varepsilon_t\}$?

(a) The elements in the sequence $\{\varepsilon_t\}$ must have no contemporaneous correlation.

(b) The elements in the sequence $\{\varepsilon_t\}$ must have no correlation across time.

(c) The vector must be independent across time.

(d) $\mathrm{E}_t[\varepsilon_t|\varepsilon_{t-1},\varepsilon_{t-2},\ldots] = 0$

(e) The vector must be conditionally homoskedastic.

16. Assuming y_t is covariance stationary and is generated by a VAR(P) with white noise residuals, which of the VAR order selection methods lead to consistent lag length selection?

(a) BIC (Bayesian)

(b) AIC (Akaike)

(c) Likelihood-ratio

(d) HQIC (Hannan-Quinn)

17. The central limit theorem ...

(a) holds with finite and asymptotic samples sizes.

(b) forms a basis for inference on the OLS regression parameters.

(c) is a distributional statement for the sample mean.

(d) states that the sample mean converges to the population mean for i.i.d. data with finite variance.

18. Which statements are true about about outliers?

    (a) A finite number of outliers result in biased OLS estimates of linear regression coefficients.
    (b) A finite number of outliers result in inconsistent OLS estimates of linear regression coefficients.
    (c) Windsorization and trimming identify outlying observations by the most extreme realizations of the dependent variable.
    (d) Trimming removes observations identified as outliers.

19. Which regression specifications can be estimated with linear regression assuming x_i is observable, $\mathrm{E}[\varepsilon_i|X] = 0$ and $\mathrm{E}[\varepsilon_i|Y_{i-1}] = 0$?

    (a) $y_i = \beta_1 x_i^{\beta_2} \varepsilon_i$, $\varepsilon_i > 0$
    (b) $y_i = \beta_1 x_i^{\beta_2} + \varepsilon_i$, $\varepsilon_i > 0$
    (c) $y_i = \sqrt{\sigma_i^2}\varepsilon_i$, with $\sigma_i^2 = \omega + \alpha y_{i-1}^2 + \beta \sigma_{i-1}^2$
    (d) $y_i = \beta_1 \sin x_i + \beta_2 \ln x_i + \varepsilon_i$, $x_i > 0$

20. Which of the following terms are sources of non-stationarity in the time-series model $y_t = \phi + \delta t + y_{t-1} + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t$, where $\varepsilon_t \overset{\text{i.i.d.}}{\sim} N(0, \sigma_\varepsilon^2)$ and $\mathbf{x}_t \overset{\text{i.i.d.}}{\sim} N(\mathbf{0}, \Sigma)$ are bivariate normally distributed?

    (a) The intercept ϕ.
    (b) The correlated regressors x_{1t} and x_{2t}.
    (c) The deterministic trend (δt).
    (d) The lag (y_{t-1}).

Part B: Long Answer

Answer THREE of the seven questions in this section.

Each question is worth 10% of the exam mark (i.e., 1/3 of 30%). Within each question points sum to 100% and so will be scaled by 10% when combined in the final exam mark. Answers must be as precise as possible, i.e., should use mathematical notation and formulae where relevant.

1. Suppose

$$\mathbf{y}_t = \begin{bmatrix} 0.2 & 0.4 \\ 0.0 & 0.6 \end{bmatrix} \mathbf{y}_{t-1} + \begin{bmatrix} 0.0 & 0.4 \\ 0.1 & 0.3 \end{bmatrix} \mathbf{y}_{t-2} + \varepsilon_t$$

   where ε_t is a vector white noise process with covariance Σ. Answer the following questions about the VAR above:

   (a) [30%] In bivariate a VAR(2), what restrictions on the model's parameters are implied if y_2 does not Granger Cause y_1?

   (b) [30%] Write the model in error correction form.

   (c) [20%] Using the coefficient matrix on $\mathbf{y}_{t-1}$ in the ECM, determine if the model is (a) a cointegrated VAR, (b) 2 random walks, or (c) covariance stationary. Note that in a 2 by 2 matrix $\mathbf{A}$, the eigenvalues are the solution to $\lambda_1\lambda_2 = a_{11}a_{22} - a_{12}a_{21}$ and $\lambda_1 + \lambda_2 = a_{11} + a_{22}$.

   (d) [20%] What are the 1-step and 2-step ahead forecast from the ECM for $\mathrm{E}_t[\mathbf{y}_{t+h}]$, $h = 1, 2$.

2. Suppose two assets, X and Y, have are bivariate normally distributed with $\mu_X = 8\%$, $\mu_Y = 5\%$, $\sigma_X^2 = 0.25^2$, $\sigma_Y^2 = 0.15^2$ and $\rho_{XY} = -0.3$.

   (a) [20%] What the expected return to a portfolio $Z = wX + (1-w)Y$?

   (b) [20%] What is the variance of Z as a function of w?

   (c) [40%] What value of w minimizes the variance of the portfolio?

   (d) [20%] What value of w maximizes the Sharpe ratio of the portfolio, $\mathrm{E}[Z]/\sqrt{\mathrm{V}[Z]}$?

3. Suppose you observe a sequence of n i.i.d. data from a Poisson(λ) distribution where each observation has pmf

$$f(x;\lambda) = \frac{\lambda^x \exp(-\lambda)}{x!}.$$

   (a) [30%] What is the MLE of $\hat{\lambda}$?

   (b) [20%] What is the asymptotic distribution of the MLE?

Use the sample

$$\{4,5,4,6,3,5,5,6,3,3\}$$

to answer the questions (c) and (d).

(c) [25%] Using the data above, test the null $H_0 : \lambda = 3.3$ using a t-test and a 5% test size. The lower-tail quantiles from a normal distribution are in the table below.

(d) [25%] Repeat the test in (c) using a Likelihood ratio test.

Quantile	Value
1%	-2.32
2.5%	-1.95
5%	-1.64
10%	-1.28

4. Answer the following two questions about causal inference:

(a) [50%] Describe 2 methods to identify a causal effect in observational (non-experimental) data. Compare the two methods and discuss their advantages and limitations.

(b) [50%] How might an RCT, which is often referred to as the gold standard for causal effect estimation, produce a misleading estimate?

5. Suppose $y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_{12} y_{t-12} + \theta \varepsilon_{t-1} + \varepsilon_t$ and $\varepsilon_t \overset{\text{i.i.d.}}{\sim} WN(0, \sigma_\varepsilon^2)$. For the parts (a) - (c) assume that parameters are consistent with y_t being a covariance stationary process. Answer the following questions:

(a) [20%] What is the value of $\mathrm{E}[y_{t+2}]$?

(b) [20%] What is the value of $\mathrm{E}_t[y_{t+2}]$?

(c) [20%] What is the value of $\lim_{h\to\infty} \mathrm{E}_t[y_{t+h}]$?

(d) [20%] Now we do not assume $y_t$to be covariance stationary. Let $\phi_0 = 8$, $\phi_1 = 0.8$, $\phi_2 = -0.15$ and $\theta = 12$. Is y_t stable for the given parameters?

(e) [10%] Rewrite the model using only differenced data Δy_t, Δy_{t-1}, $\Delta y_{t-2} \ldots$ and y_{t-1}?

(f) [10%] How would you describe the process $\{y_t\}$ if the coefficient on y_{t-1} is 0 in this form?

6. If $\ln RV_t$ is modeled as a HAR

$$\ln RV_t = 0.1 + 0.4 \ln RV_{t-1} + 0.3 \ln RV_{t-1:5} + 0.22 \ln RV_{t-1:22} + \varepsilon_t$$

where $\varepsilon_t \sim N(0, \sigma^2)$ where $\ln RV_{t-1:h} = h^{-1} \sum_{i=1}^{h} \ln RV_{t-i}$ is the average of h lags of $\ln RV$.

(a) [20%] What is $\mathrm{E}_t[\ln RV_{t+1}]$?

(b) [20%] What is $\mathrm{E}_t[\ln RV_{t+2}]$?

(c) [20%] What is $\lim_{h\to\infty} \mathrm{E}_t[\ln RV_{t+h}]$?

(d) [20%] What is the conditional distribution of the 2-step forecast error, $\ln RV_{t+2} - \mathrm{E}_t\left[\ln RV_{t+2}\right]$?

(e) [10%] What is $\mathrm{E}_t\left[RV_{t+1}\right]$?

(f) [10%] What is $\mathrm{E}_t\left[RV_{t+2}\right]$?

7. Suppose the correct model is

$$y_i = x_{1,i}\beta_1 + x_{2,i}\beta_2 + \varepsilon_i, \tag{1}$$

where $i = 1,...,n$ and the researcher estimates

$$y_i = x_{1,i}\beta_1 + v_i.$$

(a) [20%] What is the effect on $\hat{\beta}_2$ of omitting the variable $x_{1,i}$ in the regression equation?

(b) [20%] Is there always a cost of missing $x_{2,i}$ in the regression equation? If not, give two examples.

(c) [20%] Suppose the researcher did not include $x_{2,i}$ because she cannot access this variable. Explain how she can use an instrumental variable z_i to fix the problematic estimate $\hat{\beta}_1$.

Now suppose the correct model is

$$y_i = x_{1,i}\beta_1 + \varepsilon_i,$$

and the researcher estimates the larger model

$$y_i = x_{1,i}\beta_1 + x_{2,i}\beta_2 + \varepsilon_i.$$

(d) [20%] What is the cost of adding an unnecessary variable $x_{2,i}$?

(e) [20%] Explain how the researcher can use cross-validation to select which variables to include in cross-sectional regression models.

Part C: Extended Answer

Answer ALL questions in this section.

The section contributes 30% towards the final mark.

1. Your colleague has built two new models for Value-at-Risk, both using magical machine learning methods (Models 2 & 3). Your colleague needs you to validate that her machine learning approach is a good alternative to Filtered Historical Simulation (Model 1). While she did not leave you the code or the raw data, she has produced some basic statistics and visualizations (Tables 1 and 2 and Figure 1). All models are fit to the same return data, and all results are out-of-sample.

   (a) [67%] Use the available measures to construct the best story you can about whether you think your firm should move to one of the machine-learning-based Value-at-Risk or remain with Filtered Historical Simulation. The best answers will use mathematical notation were relevant and compute statistics using the data in the tables when these can be transformed into measures of absolute or relative performance of the models.

   (b) [33%] Why do we use the tick-loss function when forecasting Value-at-Risk? Explain how the tick-loss function is like the Mean Square Error (MSE) loss function that is used when foresting the conditional mean or the Quasi-likelihood-loss (QLIK) function that is used when forecasting the conditional variance.

Statistics Computed using the HITs

	Summary Statistics		
	Model 1	Model 2	Model 3
$\hat{\mu}$	0.0111	-0.0126	0.0164
$\hat{\sigma}$	0.3144	0.2824	0.3209
$\hat{\sigma}_{NW}$	0.2964	0.3831	0.3475
T	756	756	756
$\sum_{t=1}^{T-1} I_{\left[r_t<-VaR_t^j\right]} I_{\left[r_{t+1}<-VaR_{t+1}^j\right]}$	7	13	28
$\sum_{t=1}^{T-1}\left(1-I_{\left[r_t<-VaR_t^j\right]}\right)\left(1-I_{\left[r_{t+1}<-VaR_{t+1}^j\right]}\right)$	594	636	607
$\text{Corr}\left[HIT_t, HIT_{t+1}\right]$	-0.03142	0.1200	0.2282

	Covariance $\left(\hat{\Sigma}\right)$		
	Model 1	Model 2	Model 3
Model 1	0.0987	0.0577	0.0743
Model 2	0.0577	0.0796	0.0506
Model 3	0.0743	0.0506	0.1028

	Long-run Covariance $\left(\hat{\Sigma}_{NW}\right)$		
	Model 1	Model 2	Model 3
Model 1	0.0878	0.0744	0.0813
Model 2	0.0744	0.1467	0.0793
Model 3	0.0813	0.0793	0.1207

Table 1: This table contains statistics based on the sequence of HITs defined as $I_{\left[r_{t+1}<-VaR_{t+1}^j\right]} - \alpha$ for models $j = 1,2,3$ where $\alpha = 10\%$. The top panel contains the mean of the HITs ($\hat{\mu}$), the standard deviation of the HITs ($\hat{\sigma}$), the long-run standard deviation of the HITs computed as the square root of a Newey-West variance using 12 lags ($\hat{\sigma}_{NW}$), the number of out-of-sample observations (T), the number of periods where a VaR violation (an exceedance) was followed by a VaR violation $\left(\sum_{t=1}^{T-1} I_{\left[r_t<-VaR_t^j\right]} I_{\left[r_{t+1}<-VaR_{t+1}^j\right]}\right)$, the number of periods where no VaR violation was followed by no VaR violation $\left(\sum_{t=1}^{T-1}\left(1-I_{\left[r_t<-VaR_t^j\right]}\right)\left(1-I_{\left[r_{t+1}<-VaR_{t+1}^j\right]}\right)\right)$, and the correlation of the HITs across two consecutive periods ($\text{Corr}\left[HIT_t, HIT_{t+1}\right]$). The middle panel contains the covariance of the HITs across the three methods ($\hat{\Sigma}$). The final panel contains the long-run covariance of the HITs measured using a Newey-West covariance estimator with 12 lags ($\hat{\Sigma}_{NW}$).

Statistics Computed using the Tick Losses

	Mean		
	Model 1	Model 2	Model 3
$\bar{L}$	0.1273	0.1712	0.1294

	Covariance $\left(\hat{\Sigma}\right)$		
	Model 1	Model 2	Model 3
Model 1	0.0426	0.0433	0.0381
Model 2	0.0433	0.0752	0.0364
Model 3	0.0381	0.0364	0.0353

	Long-run Covariance $\left(\hat{\Sigma}_{NW}\right)$		
	Model 1	Model 2	Model 3
Model 1	0.1591	0.1783	0.1543
Model 2	0.1783	0.2332	0.1702
Model 3	0.1543	0.1702	0.1513

Table 2: The top panel contains the mean tick-loss for each of the models computed using $\alpha = 10\%$. The middle panel contains the covariance $\left(\hat{\Sigma}\right)$ of the tick-losses estimated using the standard covariance estimator. The bottom panel contains an estimate of the long-run covariance $\left(\hat{\Sigma}_{NW}\right)$ of the tick-losses estimated using a Newey-West covariance estimator and 12 lags.

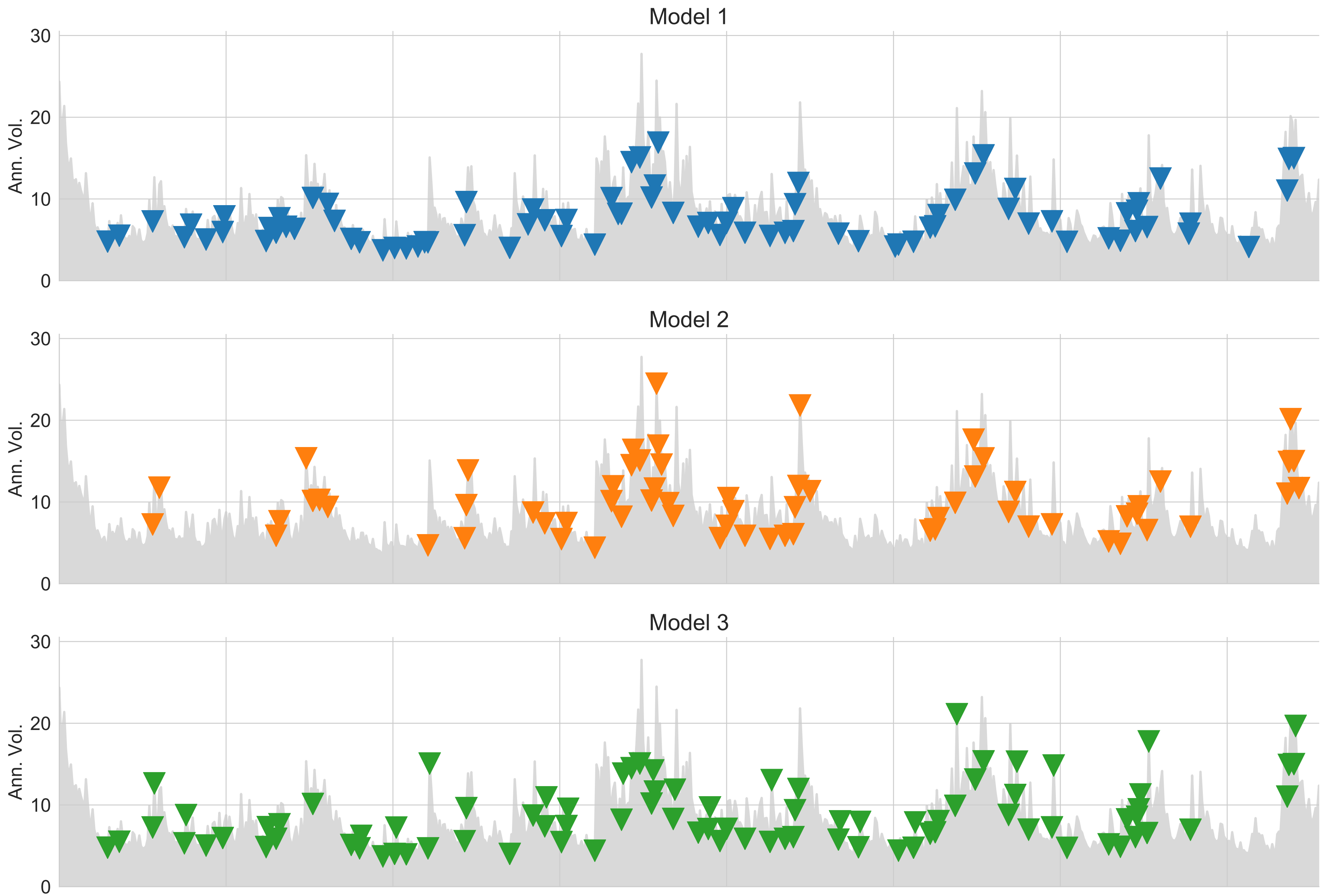

Figure 1: Plots of the VaR violations for each of the three models (▼) along with the fitted volatility, which is the same in all three panels since the underlying asset is identical.